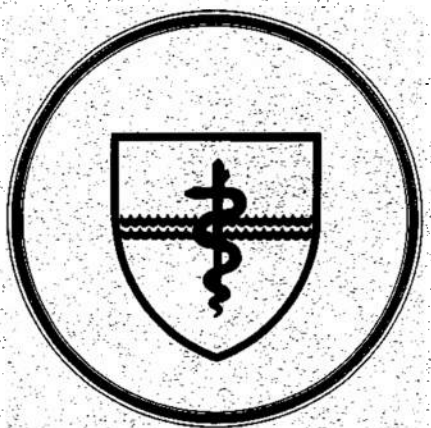
NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

SUBMARINE BASE, GROTON, CONN.



REPORT NUMBER 926

ON THE USE OF A THREE-WORDS-PER-ITEM FORMAT IN TESTS FOR THE HEARING OF SPEECH

by

J. Donald Harris

Naval Medical Research and Development Command
Research Work Unit MF58.524.004-9024

Released by:

R. A. Margulies, CDR, MC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory

18 February 1980

SUMMARY PAGE

THE PROBLEM

To evaluate recent developments in tests for the hearing of speech.

FINDINGS

Three-monosyllable-per-item tests with closed-response sets are significant advances over earlier tests using single monosyllable items or sentences in test methodology.

APPLICATION

The results contribute to the development of methods for assessing auditory performance for submarine and shipboard duty.

ADMINISTRATIVE INFORMATION

This investigation was conducted as part of Naval Medical Research and Development Command Research Work Unit Number MF58.524.004-9024 - "Determination of new minimum auditory standards for submarine duty: entrance, retention, waiver, and disqualification levels." The present report was submitted for review in August 1979, approved for publication in Sept 1979, and designated as NavSubMedRschLab Report No. 926. It appeared in the Journal of the Acoustical Society of America 67(1), in January 1980.

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

On the use of a three-words-per-item format in tests for the hearing of speech[a)]

J. Donald Harris

U.S.N. Submarine Medical Research Laboratory, Groton, Connecticut 06340
(Received 14 February 1979; accepted for publication 12 September 1979)

Single-word lists and sentence lists each have their own advantages and disadvantages for testing hearing for speech. A short history is offered of the attempts since 1941 to achieve the advantages of sentential material by presenting strings of grammatically unrelated words. Such material retains the several advantages of single-word tests. At least two recent tests using a three-monosyllables-per-item format with closed-response sets are seen to make significant advances over earlier tests in maximizing advantages and minimizing disadvantages of material presented.

PACS numbers: 43.70.Ep, 43.70.Dn

In pronouncing single words in isolation as a test for the hearing of speech, some workers have expressed dissatisfaction with (a) the time it takes to obtain responses to a relatively few stimuli and (b) the loss of all the acoustic cues of prosody and intonation and of simpler transitions in going from one word to the next. Haagen (1945) noted that word order, speaking rate, and phrasing are variables not incorporated easily into single-word-per-item tests. Furthermore, he pointed to a "set" toward context which could aid in perception. Harris (1960) noted that brief verbal stimuli cannot well be used to study many of the important types of distortion found in everyday life. For example, it was said to make no sense to reverberate a single syllable, since, before the reverberation could have any effect, the information would already have been transmitted. Also, one cannot easily quantify the speedup of very brief syllables, nor equate interruption cycles from syllable to syllable.

Unwanted complexities often are added to a test when single-word lists are abandoned in favor of linguistically meaningful sentences. This point has been made many times. Efforts have been expended to secure the advantages of sentential material with much of the linguistic cueing removed. MacFarlan (1945) proposed lists of "Nonsense Sentences" (e.g., "Scissors cut holes in clouds," "You cannot write with a hammer"), where the subject had to repeat the exact words. Speaks and Jerger (1965) constructed nonsense sentences using strings of seven words representing third-order approximations to actual English sentences (e.g., "Down by the time is real enough").

Linguistic cues cannot be avoided easily when using sentences, even with MacFarlan's or Speaks and Jerger's techniques. In addition, subjects' responses are difficult to score and interpret. In this regard, single-word tests are far superior in that write-down answers can be avoided altogether by using a closed set of possible responses. Such sets, properly constructed, can lead to phoneme confusion matrices and analyses of fine phoneme discriminations going far beyond Fletcher's (1929) concept of intelligibility as the "percent of ideas expressed in the form of simple test sentences which, after transmission, are correctly understood" (p. 264).

A compromise was advanced by Berger (1969) in which a single-word discrimination test is embedded in actual sentences. The subject listens to the sentence and thus has Haagen's (1945) "set" toward context, but the context does not allow distinguishing among the choices given. It is assumed that all five choice words are equally probable in each such sentence as the following:

weeds

seeds

"We found some wheels in the yard."

reeds

beads

The subject underlines whichever one of the five choices he understood the talker to say.

Watson and Knudsen (1940) first moved from a single-word to a multiple-word presentation without linguistic content by having the talker utter an introductory carrier phrase and a string of three key words (e.g., "The first is bait, set, ret;" "Listen to bite, rim, let;" "Try to hear beak, fle, wish"). The phonemes underlined were the only ones scored. Watson and Knudsen constructed phonograph records of three-words-per-item tests with directions for assessing a patient's speech reception threshold. They credited L. W. Sepmeyer with constructing a list of 69 words, each suitable for examining reception of a particular English phoneme. For each item, the subject wrote down all key words heard. This is not a test of speech discrimination among words within an item; it is simply a quick way to present 75 words in one session.

This test would allow confusion matrices to be drawn by noting the phoneme written as compared with the target phoneme in each word, but it would be a laborious process and relatively inexact with its open-ended response possibilities. These disks were used by Watson and Knudsen in an extensive study of selective amplification for hearing aid wearers. Unfortunately, normative data were not published nor were the 16 permutations of words or the extremely valuable phonograph records ever released.

[a)] The views expressed here are not necessarily the official position of the U. S. Navy.

Haagen (1945) reported a group of quick three-words-per-item, multiple-choice intelligibility tests. Words were drawn from a pool of 1200 one- or two-syllable words judged to be within the vocabulary of high school sophomores. These were listened to by 240–300 men, and the most frequent substitutions were used as foils in the final multiple-choice format. In Form A, for example, Talker 1 read eight items:

Item 1. swarm canvas quart
,,
,,
Item 8. knuckle dress screech

and the subject underlined on his answer sheet the words he heard:

Item 1.	form	campus	court
	warm	canvas	fort
	swarm	pamphlet	port
	storm	panther	quart
,,			
,,			
Item 8.	uncle	dread	screech
	buckle	dress	preach
	knuckle	rest	reach
	stucco	red	street

Talkers Nos. 2–12 each read eight similarly constructed items in order, for a total of 96 items (8 items for each of 12 talkers). Form B contained 96 similar items.

Haagen pointed out that a multiple-choice test, even of single words, can present about twice as many words per unit time as a write-down format, and that the use of a three-words-per-item format can reduce testing time by an additional one-half to two-thirds. Haagen's test is a quick way to administer any number of common words at any level (s) desired and can be adapted for machine scoring, but does not make it possible really to analyze errors. The full lists were printed, but never recorded.

Versions of multiple-word testing have recently been introduced with all the virtues of economy of Haagen's test and in which the most precise error analyses are made possible, Three monosyllables are pronounced as a string, with little or no linguistic connection, and closed-response sets are provided. An attempt is made with these tests to secure the following advantages of single-word tests: (a) easy group administration, (b) machine scoring, (c) the incorporation of the finest cues for phonetic discrimination, and (d) the creating and hyperfine analysis by computer of confusion matrices. At the same time they are designed to retain by a sentencelike utterance of real words the "set" of a subject expecting some sort of context and particularly to introduce all the qualities of perceptual cueing inherent in naturalistic phrasing, prosody, and coarticulation among adjacent words. As an additional goal, they are designed to minimize the linguistic cues ordinarily available.

Williams *et al.* (1976) developed a three-words-per-item test based upon the lists of the Modified Rhyme Test (MRT) (House *et al.*, 1965) which incorporates the acoustic features of sentential material, spoken and scored as the test of Haagen but allowing for all the powerful analyses of which the MRT is capable. Sergeant *et al.* (1979) have done the same but used the even more difficult discriminations of Griffiths' (1967) Diagnostic Articulation Test (DAT). These tests easily may come to displace their less efficient originals for the purposes of assessing communications efficiency either of a circuit or of an individual subject or patient. They illustrate well how by progressive stages a good original idea can be explored in depth to produce more and more powerful test instruments.

A feature of the DAT, as compared with the MRT, is that the DAT needs less degradation for use with normal talkers and listeners. With the MRT, for example, normal performance in quiet is 98% correct or better, so that it is necessary to add noise at really quite an unfavorable speech/noise ratio so that performance is reduced enough to avoid the "ceiling" effect. It is necessary with the DAT, because of its more difficult discriminations, to add noise at a significantly less intense level to achieve the same loss in performance. Now to the extent that adding noise changes the essential nature of a speech discrimination task, this difference between the MRT and the DAT is to the advantage of the DAT.

Suppose one wished to examine the effect of introducing controlled amounts of reverberation into a circuit. To be forced to introduce noise also (because otherwise the ceiling effect would render it impossible to uncover effects of slight reverberation) might well lead to an interaction between noise and reverberation which would obscure the exact effects of reverberation *per se*. In such cases it would be wise to introduce as little noise or any other degradation as possible.

The mating of the three-words-per-item format with lists of monosyllables, in particular the DAT, renders possible, really for the first time, a spectrum of experiments with respect both to talkers and to listeners. I suggest here some studies on talkers: comparing the enunciation of a word uttered in isolation versus in the middle of a three-word "sentence," with the use of either trained or naive listeners, or with computer analyses of the acoustics of the utterances:

(1) The experimenter could determine quite precisely how the enunciation of the medial phonemes, for any talker, varied with the particular phonemes of the initial and final words of the "sentence."

(2) Developmental schedules could be constructed for children of all ages, as easily as they could be induced to utter three-word strings, for the emergence of the adult form of the transitions from one phoneme to another in normal conversation, and it could be determined how these normal forms are eroded with the aging process.

(3) In children with true speech pathologies, as well as those who are simply a little late in forming the /s/,

a determination in some cases could be made of which phoneme-to-phoneme transitions were most affected, and a precise quantification made of the changes toward normal which were the result of speech therapy.

(4) In talkers who are demonstrably more intelligible in noise, or in some particular type of noise, or whose intelligibility holds up well when the sample is speeded up, or interrupted at various duty cycles, etc., a careful analysis of individual differences in manner transition from the phoneme to another, or one class of phonemes to another class, might reveal essential differences among talkers.

(5) One may study the change in pattern of consonant confusions for CV vs VC vs CC clusters by position of occurrence (initial or final) in a syllable.

There are also possible studies on listeners: comparing the intelligibility, to a particular listener or to a class of listeners, of words uttered in isolation versus in the middle of a three-word sentence:

(1) The experimenter could determine the sensitivities of the listener (s), or of the communication circuit components, or the slightest meaningful difference among phonemes, and the experimenter could determine how, within a class of phonemes most easily confused when uttered in isolation, listener (s) can or cannot detect the slight and even untranscribable variations introduced within phonemes when embedded in sentential material.

(2)–(3) Developmental and aging schedules, and quantification of listening strategies and abnormalities, could be determined for those who could be induced to respond with the three-words-per-item format.

(4) In listeners, or classes of listeners, for whom speech discrimination is noticeably good (or poor) under, for example, picket-fence noise masking but is relatively better (or worse) under, let us say, low-pass frequency filtering, a study of the confusions among particular phonemes or phoneme transitions might specify precisely the conditions for those listeners under which particular speech units are perceived well or ill.

In a society which puts so much emphasis on hearing the spoken voice rather than on reading the printed word, probably a sharply expanded effort in documenting the characteristics of a person's voice and of his hearing skills throughout life might be well received, and not only voiceprints and audiograms, but also detailed analyses of both voice and hearing, might be fondly filed in family albums along with the snapshots.

It might be thought that the discriminations suggested here are hyperfine and of little practical consequence. But in my observations of the "speech and hearing therapist" in our public schools and neighborhood clinics, and in my readings in speech and hearing journals, I find that developmental schedules and norms for either speaking or listening are of the coarsest. Objective testing of the success (or failure) of "speech and hearing therapy" by a disinterested third party, as required by any reasonable accountability program, is quite unknown other than perhaps a general assessment of a child on a three-point scale ("shows no progress," "shows progress," "shows good progress"). The usual clinician has really an imprecise grasp of the speaking or hearing abilities of the client and can describe in only the grossest terms the outcome of a regimen. Fine-tuned tests to document stages of progress in an acceptable manner are only now being constructed.

Berger, K. W. (1969). "A Speech Discrimination Task Using Multiple-Choice Key Words in Sentences," J. Aud. Res. 19, 247–262.

Fletcher, H. (1929). *Speech and Hearing* (Van Nostrand, New York).

Griffiths, J. D. (1967). "Rhyming Minimal Contrasts: A Simplified Diagnostic Articulation Test," J. Acoust. Soc. Am. 42, 236–241.

Haagen, C. H. (1945). "Intelligibility Measurement: Twenty-Four Multiple-Choice Tests," Office of Scientific Research and Development, Report No. 5567, U. S. Department of Commerce PB 12050.

Harris, J. D. (1960). "Combinations of Distortion in Speech," Arch. Otolaryngol. 72, 227–232.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (1965). "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set," J. Acoust. Soc. Am. 37, 158–166.

MacFarlan, D. (1945). "Speech Hearing Tests," Laryngoscope 55, 71–115.

Sergeant, R. L., Atkinson, J. E., and Lacroix, P. G. (1979). "The NSMRL Tri-Word Test of Intelligibility (TTI)," J. Acoust. Soc. Am. 65, 218–221.

Speaks, C., and Jerger, J. (1965), "Method for Measurement of Sentence Identification," J. Speech Hear. Res. 8, 185–194.

Watson, N. A., and Knudsen, V. O. (1940). "Selective Amplification in Hearing Aids," J. Acoust. Soc. Am. 11, 406–419.

Williams, C. E., Mosko, J. D., and Greene, J. W. (1976). "Evaluating the Ability of Aircrew Personnel to Hear Speech in Their Operational Environments," Aviat. Space Environ. Med. 47, 154–158.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NSMRL Report No. 926	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* ON THE USE OF A THREE-WORDS-PER-ITEM FORMAT IN TESTS FOR THE HEARING OF SPEECH		5. TYPE OF REPORT & PERIOD COVERED Interim report 6. PERFORMING ORG. REPORT NUMBER NSMRL Report No. 926
7. AUTHOR(s) J. Donald HARRIS		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Submarine Medical Research Laboratory Box 900 Naval Submarine Base Groton, Connecticut 06340		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MF58.524.004-9024
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Submarine Medical Research Laboratory Box 900 Naval Submarine Base Groton, Connecticut 06340		12. REPORT DATE 19 February 1980 13. NUMBER OF PAGES 3
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Naval Medical Research and Development Command National Naval Medical Center Bethesda, Maryland 20014		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Speech audiometry; Auditory standards in submarines; Hearing loss; Physical defects.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Single-word lists and sentence lists each have their own advantages and disadvantages for testing hearing for speech. A short history is offered of the attempts since 1941 to achieve the advantage of sentential material by presenting strings of grammatically unrelated words. Such material retains the several advantages of single-word tests. At least two recent tests using a three-monosyllables-per-item format with closed-response sets are seen to make significant advances over earlier tests in maximizing advantages

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

and minimizing disadvantages of material presented.